

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

THE NEW YORK TIMES COMPANY,

Plaintiff,

v.

MICROSOFT CORPORATION, OPENAI, INC.,
OPENAI LP, OPENAI GP, LLC, OPENAI, LLC,
OPENAI OPCO LLC, OPENAI GLOBAL LLC,
OAI CORPORATION, LLC, and OPENAI
HOLDINGS, LLC,

Defendants.

Civil Action No. 1:23-cv-11195-SHS

**PLAINTIFF'S RESPONSE TO DEFENDANT OPENAI OPCO, LLC'S OBJECTION TO
NONDISPOSITIVE ORDER PURSUANT TO FEDERAL RULE OF CIVIL
PROCEDURE 72(a) (DKT. 362)**

TABLE OF CONTENTS

INTRODUCTION	1
BACKGROUND	2
I. The Times’s Complaint Alleges Harm from Defendants’ Unlicensed Use of The Times’s Content and the Infringing Outputs of Defendants’ Tools.....	2
II. Magistrate Judge Wang’s Well-Reasoned November 22, 2024 Order was Consistent with Governing Law.....	3
III. The Times Already Agreed to Produce a Substantial Number of Documents About its Use of Defendants’ Generative AI Products.	5
LEGAL STANDARD	6
ARGUMENT	7
I. The Court Properly Considered the Statutory Fair Use Factors.	7
II. OpenAI’s Requests are Not Relevant to the Market Harm Inquiry.	8
III. OpenAI’s Requests are Not Relevant to Assessing the Transformativeness or Public Benefit of OpenAI’s Unauthorized Uses of Times Content.....	14
IV. OpenAI’s Requests are Not Relevant to Damages.	15
CONCLUSION.....	17

TABLE OF AUTHORITIES

	Page(s)
Cases	
<i>Am. Geophysical Union v. Texaco Inc.</i> , 60 F.3d 913 (2d Cir. 1994).....	10
<i>Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith</i> , 598 U.S. 508 (2023).....	3, 15
<i>Castle Rock Entm’t v. Carol Pub. Grp., Inc.</i> , 955 F. Supp. 260 (S.D.N.Y. 1997)	16
<i>Coventry Cap. US LLC v. EEA Life Settlements Inc.</i> , 333 F.R.D. 60 (S.D.N.Y. 2019)	6
<i>Fed. Trade Commission v. BlueHippo Funding, LLC</i> , 2016 WL 10571896 (S.D.N.Y. Apr. 19, 2016).....	16
<i>Google LLC v. Oracle Am., Inc.</i> , 593 U.S. 1 (2021).....	<i>passim</i>
<i>Hachette Book Grp., Inc. v. Internet Archive</i> , 115 F.4th 163 (2d Cir. 2024)	<i>passim</i>
<i>Kumaran v. Vision Fin. Markets, LLC</i> , 2022 WL 17540669 (S.D.N.Y. Dec. 6, 2022)	7
<i>Madi v. United Parcel Serv. Gen. Serv. Co.</i> , 1999 WL 193403 (S.D.N.Y. Apr. 7, 1999).....	6
<i>MCA, Inc. v. Wilson</i> , 677 F.2d 180 (2d Cir. 1981).....	14
<i>Nikkal Indus., Ltd. v. Salton, Inc.</i> , 689 F. Supp. 187 (S.D.N.Y. 1988)	6
<i>Pearson Educ., Inc. v. Chegg, Inc.</i> , 2023 WL 3775317 (D.N.J. June 2, 2023).....	16
<i>R.F.M.A.S., Inc. v. So</i> , 748 F. Supp. 2d 244 (S.D.N.Y. 2010).....	6, 7
<i>Sec. Pac. Mortg. & Real Est. Servs., Inc. v. Canadian Land Co. of Am., N.V.</i> , 690 F. Supp. 1214 (S.D.N.Y. 1988).....	16

<i>Shand v. Kijakazi</i> , 2023 WL 8379173 (S.D.N.Y. Dec. 4, 2023)	13
<i>United States v. Gomez</i> , 877 F.3d 76 (2d Cir. 2017).....	9
<i>Winfield v. City of New York</i> , 2017 WL 5054727 (S.D.N.Y. Nov. 2, 2017).....	14
<i>Winfield v. City of New York</i> , 2018 WL 840085 (S.D.N.Y. Feb. 12, 2018).....	15
Statutes	
17 U.S.C. § 107(1)	7, 8, 14
17 U.S.C. § 107(4)	9
Other Authorities	
Fed. R. Civ. Proc. 72(a)	1, 6, 8

Plaintiff The New York Times Company (“The Times”) respectfully submits this opposition to OpenAI’s request under Federal Rule of Civil Procedure 72(a) to modify or set aside Magistrate Judge Wang’s November 22, 2024 Order. *See* Dkt. 344. As set forth below, Judge Wang’s well-reasoned decision should be affirmed.

INTRODUCTION

From the inception of this litigation, OpenAI has mischaracterized The Times’s lawsuit as an indictment of generative AI generally, seeking to distract from its own massive copyright infringement by accusing other companies of doing the same thing. But there are only two defendants in this case, Microsoft and OpenAI, and the question of whether *other* companies engaged in illegal copying is simply not one that can be answered here. Nor is The Times “speaking out of both sides of its mouth” by using generative AI technology, as OpenAI contends, Dkt. 363 at 1, because The Times does not use other people’s copyrighted content to build large-language models (“LLMs”), *see* Exhibit (“Ex.”) 1 ¶ 4 (Affidavit of Karen Chesley). To the extent that The Times develops generative AI tools, it has done so using LLMs created by others, and pays the fees those developers—including OpenAI—charge to do so. *See* Ex. 1 ¶ 6. And where Defendants’ technology is used to aid Times reporting or its presentation of content, The Times has already agreed to produce documents to OpenAI, along with a host of other documents relating to its generative AI policies and use of AI for journalism.

What Defendants seek here is much broader: discovery into generative AI generally, regardless of which company developed it or why, along with anything related to “AI” or “copyright.” Dkt. 236-1 at 11. As Magistrate Judge Wang recognized during oral argument, such a request would encompass tools like autocorrect and spell check. *See* September 12, 2024 Hrg. Transcript at 112:21-22. Moreover, as a news publisher, The Times’s entire business relates to

“copyright” because that is the means through which The Times protects and monetizes the works it creates. Nonetheless, according to Defendants, such discovery is necessary because any benefits from *any* generative AI “technology” could support its fair use defense. *See* Dkt. 363 at 1; Dkt. 236 at 1. Judge Wang correctly held that this is wrong. The fair use inquiry looks not at benefits from a general technology, but rather at any potential benefits The Times may be reaping from Defendants’ *specific use* of copyrighted content—in this case, Defendants’ copying of The Times’s works to train and operate their generative AI tools. Put differently, even if OpenAI could prove that The Times is benefitting from the use of a Google generative AI tool, that would have no bearing on *OpenAI’s* fair use defense. Accordingly, only “benefits” that The Times receives from *Defendants’* tools could ever possibly be relevant, and The Times has agreed to produce responsive documents regarding its use of those tools for its journalism. The further discovery sought by OpenAI is neither relevant nor proportional to OpenAI’s fair use defense.

Because OpenAI has not met its exacting burden of demonstrating that Judge Wang’s decision is clearly erroneous or contrary to law, its Objection should be denied.

BACKGROUND

I. The Times’s Complaint Alleges Harm from Defendants’ Unlicensed Use of The Times’s Content and the Infringing Outputs of Defendants’ Tools.

At issue in this case are Defendants’ generative AI products, including OpenAI’s ChatGPT tool and Microsoft’s Bing Chat (now “Copilot”). The Times alleges that Defendants build and operate their generative AI tools by copying millions of The Times’s copyrighted works without The Times’s consent and without compensation. *See* Dkt. 170 at 2. Specifically, The Times’s allegations focus on the harm from Defendants’ actions: their copying of The Times’s content during training, the synthetic search process using retrieval augmented generation (“RAG”), and the generation of outputs in response to user prompts. *See id.* The Complaint is not, as OpenAI

suggests, an indictment of generative AI technology generally. There are other lawsuits pending against different generative AI developers that will address any liability they may have for copyright violations, but this is not one of them.

II. Magistrate Judge Wang’s Well-Reasoned November 22, 2024 Order was Consistent with Governing Law.

Defendants have collectively served 359 requests for production on The Times, many of which were facially overbroad and irrelevant. When The Times objected, OpenAI filed a series of motions to compel, seeking—among other things—privileged documents about The Times’s infringement investigation and all reporters’ files for the 10 million Times works that OpenAI copied. *See* Dkts. 124, 152. Judge Wang carefully reviewed OpenAI’s motions and the parties’ arguments, and denied each motion. Undeterred, on September 3, 2024, OpenAI sought to compel production of The Times’s use of nonparties’ generative AI tools, The Times’s creation and use of its own generative AI products, and all Times communications regarding “copyright” and “AI.” *See* Dkt. 236 at 1.

On November 22, 2024, the Court denied OpenAI’s motion to compel, holding it “failed to demonstrate the relevance of the information sought.” Dkt. 344 at 6. In reaching this conclusion, the Court first outlined the four statutory fair use factors and correctly noted that each factor “requires scrutiny of a *defendant’s* purported use of the copyrighted work(s), and whether that *defendant’s* use may constitute ‘fair use’ under the Act.” *Id.* at 2; *see also Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 527–28 (2023) (noting that in assessing the first statutory fair use factor, a court considers “*the copier’s* use of an original work” (emphasis added)); *Google LLC v. Oracle Am., Inc.*, 593 U.S. 1, 35 (2021) (noting that in assessing the fourth statutory fair use factor, a court “must take into account the public benefits *the copying* will likely produce” (emphasis added)). In other words, the Court reasoned that the fair use factors did not require it to

examine the entire generative AI industry, whether The Times had used tools from that industry, or the effect those tools had on The Times’s business practices. *See* Dkt. 344 at 2–3, 5. The Court noted that its conclusion was bolstered by the fact that The Times’s Amended Complaint was “tightly focused on Defendant’s particular Gen AI products and [its] alleged use of The Times’s copyrighted content,” rather than alleging a “wholesale indictment of Gen AI tools,” as Defendant repeatedly claimed. *Id.* at 3.

In so holding, the Court rejected OpenAI’s reading of *Google LLC v. Oracle America, Inc.*, 593 U.S. 1 (2021). *See* Dkt. 344 at 4. OpenAI contended that *Google* was “binding” precedent that supported its position that “[e]vidence regarding [T]he Times’s creation, use, and positions on generative AI generally is directly relevant” to the first and fourth statutory fair use factors because, they claimed, analyzing other parties’ generative AI tools would somehow show that OpenAI’s technology “serves the public interest generally.” Dkt. 236 at 1. However, the Court correctly noted that “the Supreme Court suggested a more nuanced view of the market effects” inquiry in the fourth fair use factor; Dkt. 344 at 4, specifically, one that requires a court to assess “the public benefits *the copying* will likely produce,” *Google*, 593 U.S. at 35 (emphasis added). And, in assessing the public benefits from the copying, the Court stated that discovery “would be directed to *the Defendant and the public benefits of its copying*, not whether nonparties’ Gen AI tools . . . serve a general public benefit.” Dkt. 344 at 4 (emphasis added). Accordingly, the Court concluded that “[w]hether nonparties’ Gen AI tools confer benefits on the journalism industry is not relevant to a determination of whether Defendant’s acts—i.e., the alleged copying involving Defendant’s Gen AI tools—constitute fair use.” *Id.* at 5.

III. The Times Already Agreed to Produce a Substantial Number of Documents About its Use of Defendants’ Generative AI Products.

OpenAI’s effort to overturn Judge Wang’s order follows its pattern of demanding discovery that is irrelevant, unduly burdensome, and wildly disproportionate to the needs of this case. OpenAI sought three categories of discovery about generative AI: (1) The Times’s use of generative AI, *see* Dkt. 236-1 at 14–15, (2) The Times’s training of its own generative AI tools, *see id.* at 15–16, and (3) The Times’s “policies and procedures” regarding generative AI, *see id.* at 7–8, 9, 11. Many of OpenAI’s requests were facially irrelevant. For example, RFP 50 seeks documents about an AI tool that was designed to examine satellite images to trace the path of a “Chinese spy balloon” spotted in 2023 for reporting purposes. *See id.* at 15–16. The AI satellite tool uses AI, not generative AI, so it does not even meet OpenAI’s overbroad relevance theory.

Notwithstanding its concerns about overbreadth and relevance, The Times attempted to compromise by agreeing to produce nonprivileged documents: (1) regarding The Times’s use of Defendants’ generative AI tools in reporting or presenting content, *see id.* at 15 (OAI RFP 49); (2) regarding The Times’s trainings and policies about using generative AI, *see id.* (OAI RFP 46); (3) relating to The Times’s A.I. Initiatives program, *see id.* (OAI RFP 47); and (4) involving communications between The Times and third parties regarding Defendants’ use of The Times’s content in their generative AI products and whether to license The Times’s works to OpenAI, *see id.* at 14–15 (RFP 44). *See* Ex. 2. The Times’s AI Initiatives program was developed to consider how The Times might leverage generative AI and other forms of machine learning for the benefit of The Times’s journalists and readers.¹ While The Times disagrees that discovery into its use of non-party generative AI tools is relevant to this case, The Times offered discovery regarding this

¹ *See* The New York Times, “Introducing the A.I. Initiatives Team,” May 9, 2024, *available at* <https://www.nytimes.com/press/introducing-the-a-i-initiatives-team/>.

program—discovery that concerns such non-party tools—in an effort to compromise with OpenAI, which at that time indicated its focus was on how generative AI tools could aid journalism.

Nevertheless, OpenAI remains dissatisfied. More specifically, OpenAI seeks to compel production of The Times’s use of nonparties’ generative AI tools, The Times’s creation and use of its own generative AI products, and all Times communications regarding “copyright” and “AI.” *See* Dkt. 236 at 1. For the Court’s convenience, The Times has attached an appendix which outlines the requests relevant to OpenAI’s motion to compel and the responsive discovery The Times has agreed to produce. *See* Ex. 2.

LEGAL STANDARD

Under Federal Rule of Civil Procedure 72(a), “a magistrate judge’s discovery ruling is a non-dispositive order which can only be overturned if it is ‘clearly erroneous or contrary to law.’” *Madi v. United Parcel Serv. Gen. Serv. Co.*, 1999 WL 193403, at *1 (S.D.N.Y. Apr. 7, 1999) (quoting Fed. R. Civ. P. 72(a)). “A finding is clearly erroneous if the reviewing court is left with the definite and firm conviction that a mistake has been committed.” *Coventry Cap. US LLC v. EEA Life Settlements Inc.*, 333 F.R.D. 60, 64 (S.D.N.Y. 2019) (quoting *R.F.M.A.S., Inc. v. So*, 748 F. Supp. 2d 244, 248 (S.D.N.Y. 2010)). “Similarly, a finding is contrary to law when it fails to apply or misapplies relevant statutes, case law, or rules of procedure.” *Id.* (quoting *R.F.M.A.S., Inc.*, 748 F. Supp. 2d at 248).

“Therefore, pursuant to these standards, a magistrate judge’s determinations on discovery matters are entitled to substantial deference.” *R.F.M.A.S., Inc.*, 748 F. Supp. 2d at 248; *see Nikkal Indus., Ltd. v. Salton, Inc.*, 689 F. Supp. 187, 189 (S.D.N.Y. 1988) (“Consistently, it has been held that a magistrate’s report resolving a discovery discourse between litigants should be afforded substantial deference and be overturned only if found to be an abuse of discretion.”). “The

magistrate judge’s findings should not be rejected merely because the court would have decided the matter differently.” *Kumaran v. Vision Fin. Markets, LLC*, 2022 WL 17540669, at *6 (S.D.N.Y. Dec. 6, 2022). “The party seeking to overturn a magistrate judge’s decision thus carries a heavy burden.” *R.F.M.A.S., Inc.*, 748 F. Supp. 2d at 248 (quoting *U2 Home Ent., Inc. v. Hong Wei Int’l Trading Inc.*, 2007 WL 2327068, at *1 (S.D.N.Y. Aug. 13, 2007)).

ARGUMENT

OpenAI has failed to satisfy its heavy burden of demonstrating that Magistrate Judge Wang’s November 22, 2024, Order was clearly erroneous or contrary to law. OpenAI’s argument is really that a broader industry analysis of *all* generative AI tools is needed to establish its fair use defense. But that is not the law, as set forth below. *First*, the Court appropriately considered the four statutory fair use factors when resolving this discovery dispute. *Second*, the Court correctly ruled that OpenAI’s discovery requests were not relevant to the market harm inquiry—that is, the fourth statutory fair use factor. And *third*, the Court properly determined that The Times’s use of, and comments about, “generative AI generally” were not relevant to assessing whether OpenAI’s unauthorized use of The Times’s works provided a public benefit or was transformative. *See* Dkt. 344 at 5. Magistrate Judge Wang’s Order should be affirmed.

I. The Court Properly Considered the Statutory Fair Use Factors.

First, OpenAI argues that the Court clearly erred in limiting its fair use analysis to “the statutory factors in Section 107” because it should instead have considered unspecified “issues beyond the statutory language in Section 107.” Dkt. 363 at 11. However, OpenAI does not offer a single controlling case in which a court went beyond the statutory factors when resolving a fair use dispute, nor does it explain what “issues” it was clear error for Judge Wang to purportedly ignore. Indeed, in *Google*—the case that OpenAI hinges much of its position on—the Supreme

Court explicitly limited its analysis to the four statutory fair use factors and rejected the parties’ invitation to consider extra-statutory factors. *See Google*, 593 U.S. at 32–33 (refusing to consider whether “bad faith has any role in a fair use analysis” and limiting its analysis to the Section 107 statutory fair use factors). Consistent with this precedent, Magistrate Judge Wang—while acknowledging that the statutory factors are “non-exclusive”—also declined to consider OpenAI’s extra-statutory arguments in determining that OpenAI’s discovery requests were not relevant to the issue of fair use. *See* Dkt. 344 at 2; *see also Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 578–94 (1994) (limiting fair use analysis to the Section 107 factors); *Hachette Book Grp., Inc. v. Internet Archive*, 115 F.4th 163 (2d Cir. 2024) (same). In light of the clear weight of authority that has resolved the fair use issue with reference to the statutory factors alone, it cannot be said that the Court clearly erred in declining to go beyond them.

II. OpenAI’s Requests Are Not Relevant to the Market Harm Inquiry.

Second, OpenAI argues that the Court clearly erred in ruling that The Times’s use of, and positions about, “generative AI generally” were not relevant to the market harm inquiry—that is, the fourth statutory fair use factor.² *See* Dkt. 363 at 13. This argument fails for several reasons.

A. OpenAI waived this argument by failing to raise it to Judge Wang.

In its September 3, 2024, submission to Magistrate Judge Wang, OpenAI primarily based its motion to compel on the “public benefits” prong of the fourth fair use factor. *See* Dkt. 236 at 1; *see also* Dkt. 373 at 9 (OpenAI’s Second Rule 72(a) Brief) (noting that in Magistrate Judge Wang’s November 22, 2024 Order, she denied “OpenAI[’]s request to compel production of documents based on the ‘public benefits’ prong of the fair use defense”). However, in its instant Rule 72(a)

² The Times has provided a substantial number of documents related to its “positions about” generative AI. The Times has agreed to produce, among other things, (1) its guidance on the use of generative AI tools, (2) any related non-privileged notes and talking points, and (3) communications with government agencies regarding generative AI—which are all directly responsive to OpenAI’s RFPs. *See* Dkt. 236-1 at 7-9, 11.

motion, OpenAI now argues, for the first time, that the requested evidence is also allegedly “relevant to the question [of] whether there is a” licensing market for the underlying data. Dkt. 363 at 14. OpenAI has forfeited this argument by not adequately raising it before Magistrate Judge Wang in its September 3, 2024 submission—or indeed, in any of its prior discovery correspondence with The Times—without providing any reason for its failure to do so. *See United States v. Gomez*, 877 F.3d 76, 95 (2d Cir. 2017) (noting that courts typically do not exercise their discretion to hear a forfeited argument when it “was ‘available to the parties below and they proffer no reason for their failure to raise the arguments below’” (quoting *Bogle-Assegai v. Connecticut*, 470 F.3d 498, 504 (2d Cir. 2006))).

B. The Times’s use of third-party generative AI products is irrelevant to the market harm inquiry.

Even if this Court overlooks OpenAI’s waiver, its argument is meritless. The fourth fair use factor assesses “the effect of the use upon the potential market for or value of the copyrighted work.” 17 U.S.C. § 107(4). In evaluating this factor, courts analyze “whether there was harm to the actual or potential markets for the copyrighted work.” *Google*, 593 U.S. at 26. More specifically, “[t]his factor ‘focuses on whether the copy brings to the marketplace a competing substitute for the original, or its derivative, so as to deprive the rights holder of significant revenues because of the likelihood that potential purchasers may to opt to acquire the copy [rather than] the original.’” *Hachette Book Grp., Inc.*, 115 F.4th at 189 (second alteration in original) (quoting *Authors Guild v. Google, Inc.*, 804 F.3d 202, 223 (2d Cir. 2015)). “[T]he ultimate burden of proving that the secondary use does not compete in the relevant market is appropriately borne by the party asserting the [fair use] defense: the secondary user.” *Id.* at 194.

Here, “the use” and “the copy” at issue are OpenAI and Microsoft’s unauthorized uses and copying of The Times’s asserted works, and “the potential market” is the market for the same

Times’s works. The Times has already agreed to produce ample evidence about that market and The Times’s licensing efforts, including: (i) licensing agreements related to the asserted works, regardless of whether those licenses relate to generative AI; (ii) documents sufficient to show all attempts to license the asserted works for training or use in generative AI; and (iii) communications related to executed and contemplated agreements to license Times works to train and operate generative AI models. *See* Dkt. 310-1 at 45–46.

The Times’s use of its own or third-party generative AI tools, built upon LLMs trained by non-parties, is entirely irrelevant to the market harm inquiry. *See Campbell*, 510 U.S. at 590 (holding that analyzing market harm requires courts to consider “whether unrestricted and widespread conduct of the sort engaged in *by the defendant* . . . would result in a substantially adverse impact on the potential market for the original” (emphasis added) (omission in original)); *Am. Geophysical Union v. Texaco Inc.*, 60 F.3d 913, 927 (2d Cir. 1994) (noting that in evaluating a defendant’s fair use claim, “the analysis under the fourth factor must focus on the effect of [the defendant’s copying] upon the potential market for or value of [the plaintiff’s] individual articles”). Such use has no bearing on the impact of Defendants’ products on the market for The Times’s works. And OpenAI does not even attempt to demonstrate how all of The Times’s communications regarding “copyright” and “AI” bear on the market harm inquiry. Indeed, they are facially overbroad and The Times has already agreed to provide extensive responsive discovery related to these topics. *See* Ex. 2.

OpenAI also contends that The Times “conceded the relevance of [such] market-related evidence” by serving “subpoenas on numerous third-party developers of large language models,” seeking discovery into their “[l]icensing agreements with content owners.” Dkt. 363 at 15 (second alteration in original). Thus, OpenAI asserts The Times “should not be permitted to withhold

evidence of its own assessment of the existence or absence of a market for licensing content for generative AI products while at the same time demanding market-related evidence from third parties.” *Id.* This is a red herring. As an initial matter, OpenAI is simply wrong that The Times is withholding this information; in fact, The Times has already agreed to provide exactly the types of licensing information it is seeking from third parties, as discussed above. This type of evidence, which specifically relates to outbound licensing of protected publisher content and the licensing market for publishers’ works, is clearly relevant, but it has nothing to do with all documents and communications relating to The Times’s internal use of, and comments about, generative AI generally.

OpenAI’s argument—that “The Times’s decision to use [non-Defendants’ generative AI] models knowing that they were trained on unlicensed data” (like The Times’s own protected content) would somehow validate Defendants’ own unauthorized copying practices—is also misguided. Dkt. 363 at 14. Because The Times relies on LLMs built by third parties, *see* Ex. 1 ¶ 5, any discovery OpenAI could seek would necessarily have to be focused on the conduct of *third party* AI developers, not The Times. This would in turn require extensive discovery into whether those companies copied Times content (and how much), how they used it to train their LLMs, whether those LLMs are prone to outputting Times content, how Times content was used by those companies to power their own synthetic search products, and the commercial nature of those companies’ uses. And, once OpenAI had this discovery, it would be asking the Court to reach factual conclusions about how that third-party misconduct affected the market for The Times’s intellectual property, all in service of an illogical argument that The Times is a hypocrite merely because it uses third-party generative AI products while seeking compensation for the misappropriation of its intellectual property to build different generative AI products. But internal

use of tools by The Times is wholly unrelated to determining whether or not widespread infringement of The Times’s intellectual property by Defendants—or anyone else—has or will harm the relevant market; namely, The Times’s ability to license its works for generative AI training. *See Hachette Book Grp., Inc.*, 115 F.4th at 189. Accordingly, The Times’s use of any such tools is simply not relevant to evaluating the effects of Defendants’ infringing tools on the market for The Times’s works.

C. *Google* does not support OpenAI’s discovery request.

Likely recognizing the futility of its position, OpenAI hinges the remainder of its argument on two passing statements in *Google v. Oracle*. In outlining the evidence the jury heard in reaching its verdict for Google, the Supreme Court acknowledged that the plaintiff had made numerous unsuccessful “efforts to move into the mobile phone market,” *Google*, 593 U.S. at 36, and that the plaintiff’s “executives thought that widespread use of the [defendant’s product] would benefit the [plaintiff] company,” *id.* at 31. Based on these two observations, OpenAI concludes that the Supreme Court endorsed the broad view that a plaintiff’s activities are always relevant to the fair use inquiry, and that view is purportedly “binding” on the Court. *See* Dkt. 363 at 15. That is incorrect.

In *Google*, the Supreme Court considered whether Google’s copying of a copyrighted computer program developed by Sun—which Oracle acquired in 2010—constituted a fair use. In its discussion of the fourth fair use factor, the Supreme Court observed that not every harm to a company’s revenue from copying cuts against fair use: rather, a court must consider “the source of the loss.” *Google*, 593 U.S. at 35. Copying that results in a market substitute, and loss flowing from that substitution, weighs heavily against fair use; but loss flowing from, for instance, a “lethal parody” that negatively impacts demand for the original by casting it in a negative light is not

cognizable under Copyright law. *Id.* (citation omitted). In this context, the Supreme Court noted that Sun had made numerous prior “efforts to move into the mobile phone market,” and that “Android was not a market substitute for [Sun’s] Java[] software.” *Google*, 593 U.S. at 36–37. Sun’s failed attempts to enter the mobile phone market bore on the question of whether the harm it suffered was due to market substitution. Here, by contrast, whether The Times has ever tried to develop an LLM for its own internal use has no bearing on the market substitution inquiry: The substitution at issue comes from Defendants’ model outputs, which are the result of their copying Times works and may closely summarize and/or repeat them verbatim. Stated differently, OpenAI’s attempt to excuse its own infringement of The Times works by baselessly pointing to The Times’s alleged infringement of others’ works is a non-starter.

The Supreme Court’s other observation—that “Sun executives thought that widespread use of the Java programming language,” including on Google’s Android platform, “would benefit the company”—considers the benefit from ***Google’s use***. *See Google*, 593 U.S. at 31. Consistent with that consideration, The Times has already agreed to produce discovery regarding the impact of ***Defendants’*** generative AI tools on The Times’s journalism and revenue. However, OpenAI seeks to go further and compel the production of materials relating to The Times’s use of non-Defendant generative AI tools which, as already explained, is irrelevant to the market harm inquiry.

OpenAI’s reliance on the *Google* Court’s observations fails for an additional reason: the Supreme Court was not ruling on whether the evidence was discoverable or admissible; thus, any language bearing on those inquiries is merely dicta. *See Shand v. Kijakazi*, 2023 WL 8379173, at *2 n.1 (S.D.N.Y. Dec. 4, 2023) (“It is well understood that ‘dicta’ are the ‘[t]he things the court said, not what it held; and only what a court holds is binding . . . in subsequent cases.’” (alteration and omission in original) (internal quotation marks omitted)). Comments made in *dicta* cannot

establish that a court’s ruling was clearly erroneous or contrary to the law. *See Winfield v. City of New York*, 2017 WL 5054727, at *3 (S.D.N.Y. Nov. 2, 2017) (concluding that the magistrate judge’s decision “was not contrary to law” where there was an “absence of any contravening **binding** authority” (emphasis added)).

Thus, the Court did not clearly err in ruling that OpenAI’s outstanding discovery requests were irrelevant to the market harm inquiry.

III. OpenAI’s Requests Are Not Relevant to Assessing the Transformativeness or Public Benefit of OpenAI’s Unauthorized Uses of Times Content.

Third, OpenAI contends that the Court clearly erred in ruling that The Times’s use of, and comments about, “generative AI generally” were not relevant to the transformative or public benefit prongs. *See* Dkt. 344 at 5. However, this argument fails for the very same reasons discussed *supra* Section II.

The issue of transformativeness is relevant to the first fair use factor, which assesses “the purpose and character of the use, including whether such use is of a commercial nature.” 17 U.S.C. § 107(1). The central question is “the extent to which **the secondary use** is transformative.” *Hachette Book Group, Inc.*, 115 F.4th at 179 (emphasis added). In making this assessment, courts consider “whether the copier’s use ‘adds something new, with a further purpose or different character, altering’ the original, infringed copyrighted work ‘with new expression, meaning or message.’” *Google*, 593 U.S. at 29 (quoting *Campbell*, 510 U.S. at 579). Similarly, when analyzing the fourth fair use factor, courts assess “the public benefits **the copying** will likely produce.” *Id.* at 35 (emphasis added). Courts typically balance “the benefit the public will derive if the use is permitted and the personal gain the copyright owner will receive if the use is denied.” *MCA, Inc. v. Wilson*, 677 F.2d 180, 183 (2d Cir. 1981).

As with the market harm inquiry, the relevant use at issue here is OpenAI’s unauthorized use and copy of The Times’s asserted works. *See Google*, 593 U.S. at 29, 35. And again, OpenAI fails to demonstrate how The Times’s use or comments about non-Defendant generative AI tools is relevant to the transformative or public benefit prongs. *See Andy Warhol Found.*, 598 U.S. at 509 (noting that the first fair use factor “considers the reasons for, and nature of, the copier’s use of an original work”); *Hachette Book Group Inc.*, 115 F.4th at 179 (noting that in assessing the first fair use factor, courts consider “the extent to which the secondary use is transformative”). Likewise, OpenAI does not (and cannot) explain how *The Times’s use* of nonparties’ generative AI tools—which may differ significantly in purpose and design from Defendants’ products—is relevant to the determination of whether *OpenAI’s secondary use* “adds something new” to The Times’s publications, *Andy Warhol Found.*, 598 U.S. at 509, or demonstrates “the public benefits [that OpenAI’s] copying will likely produce, *Google*, 593 U.S. at 35. And, given that OpenAI “bears the initial burden of demonstrating relevance and proportionality,” its failure to articulate an adequate theory of relevance is fatal to its request. *Winfield v. City of New York*, 2018 WL 840085, at *3 (S.D.N.Y. Feb. 12, 2018).

To the extent that OpenAI claims that The Times’s use of, or comments about, *OpenAI’s* generative AI tools might be relevant to the transformative or public benefits prong, The Times has already agreed to produce such evidence. *See* Ex. 2. Therefore, “[b]ecause Defendant has not demonstrated the relevance of the [remainder of the] information sought,” the Court did not clearly err in denying OpenAI’s motion to compel. Dkt. 344 at 2.

IV. OpenAI’s Requests Are Not Relevant to Damages.

Fourth, OpenAI argues that its outstanding discovery requests are relevant to establish the “objective reasonableness” of its conduct and its reliance on the fair use doctrine, which allegedly bear on The Times’s entitlement to damages. Dkt. 363 at 20. The Times has already agreed to

produce non-privileged documents and communications relating to any *relevant* internal analyses, discussions, or evaluations regarding whether OpenAI’s alleged use may constitute fair use—viz., the very evidence that OpenAI seeks to establish the “objective reasonableness” of its conduct. *See Pearson Educ., Inc. v. Chegg, Inc.*, 2023 WL 3775317, at *7 (D.N.J. June 2, 2023) (“Under these circumstances, discovery into whether [plaintiff] believed that [defendant’s] solutions were non-infringing under industry custom and practice is directly relevant to [defendant’s] fair use defense.”). And given that The Times has not developed its own LLMs, OpenAI’s assertion that The Times relied on the very same fair use defense is complete fiction. Ex. 1 ¶¶ 3, 5. Its attempt to seek discovery to show otherwise is nothing more than “pure speculation that amount[s] to nothing more than a fishing expedition.” *Fed. Trade Commission v. BlueHippo Funding, LLC*, 2016 WL 10571896, at *1 (S.D.N.Y. Apr. 19, 2016) (citation omitted).

Moreover, as discussed *supra*, none of OpenAI’s outstanding requests are relevant to its fair use defense. Given that the requested evidence cannot, as a matter of law, support OpenAI’s affirmative defense, it follows that it cannot inform Defendants’ alleged reliance on the doctrine. *Cf. Castle Rock Entm’t v. Carol Pub. Grp., Inc.*, 955 F. Supp. 260, 267 (S.D.N.Y. 1997) (finding willfulness where the defendants had constructive notice of the infringement because they had experience with the copyright laws and the work included copyright information, as well as actual notice of the infringement but continued to publish and distribute the work); *Sec. Pac. Mortg. & Real Est. Servs., Inc. v. Canadian Land Co. of Am., N.V.*, 690 F. Supp. 1214, 1225 (S.D.N.Y. 1988) (“[E]ach of the affirmative defenses sought to be supported by discovery is deficient as a matter of law. No amount of discovery could establish a valid defense on any of the grounds asserted.”). Thus, the Court did not clearly err in denying OpenAI’s motion to compel on this basis.

CONCLUSION

For the foregoing reasons, The Times respectfully requests the Court affirm Magistrate Judge Wang's November 22, 2024 Order.

Dated: January 17, 2025

/s/ Ian Crosby

Ian Crosby (*admitted pro hac vice*)
Genevieve Vose Wallace (*admitted pro hac vice*)
Katherine M. Peaslee (*admitted pro hac vice*)
SUSMAN GODFREY L.L.P.
401 Union Street, Suite 3000
Seattle, WA 98101
Telephone: (206) 516-3880
Facsimile: (206) 516-3883
icrosby@susmangodfrey.com
gwallace@susmangodfrey.com
kpeaslee@susmangodfrey.com

Davida Brook (*admitted pro hac vice*)
Emily K. Cronin (*admitted pro hac vice*)
Ellie Dupler (*admitted pro hac vice*)
SUSMAN GODFREY L.L.P.
1900 Ave of the Stars, Suite 1400
Los Angeles, CA 90067
Telephone: (310) 789-3100
Facsimile: (310) 789-3150
dbrook@susmangodfrey.com
ecronin@susmangodfrey.com
edupler@susmangodfrey.com

Elisha Barron (5036850)
Zachary B. Savage (ZS2668)
Tamar Lusztig (5125174)
Alexander Frawley (5564539)
Eudokia Spanos (5021381)
SUSMAN GODFREY L.L.P.
1301 Avenue of the Americas, 32nd Floor
New York, NY 10019
Telephone: (212) 336-8330
Facsimile: (212) 336-8340
ebarron@susmangodfrey.com
zsavage@susmangodfrey.com
tlusztig@susmangodfrey.com

afrawley@susmangodfrey.com
espanos@susmangodfrey.com

Scarlett Collings
SUSMAN GODFREY L.L.P.
1000 Louisiana, Suite 5100
Houston, TX 77002
Telephone: (713) 651-9366
Facsimile (713) 654-6666
scollings@susmangodfrey.com

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Kristen J. Logan (*admitted pro hac vice*)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
901 New York Avenue, N.W., Suite 900 East
Washington, DC 20001
Telephone: (202) 783-6040
Facsimile: (202) 783 6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com
klogan@rothwellfigg.com

Attorneys for Plaintiff
The New York Times Company